

Generator Handler

RunPod provides a robust streaming feature that enables users to receive real-time updates on job outputs, mainly when dealing with Language Model tasks. We support two types of streaming generator functions: regular generator and async generator.


```
import runpod

def generator_handler(job):
    for count in range(3):
        result = f"This is the {count} generated output."
        yield result

runpod.serverless.start(
    {
        "handler": generator_handler, # Required
        "return_aggregate_stream": True, # Optional, results available via /run
    }
)
```

Return aggregate Stream

By default, when a generator handler is running, the fractional outputs will only be available at the `/stream` endpoint, if you would also like the outputs to be available from the `/run` and `/runsync` endpoints you will need to set `return_aggregate_stream` to True when starting your handler.

 [Edit this page](#)